

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	34576
ToLID	xcNauMacr1
Species	Nautilus macromphalus
Class	Cephalopoda
Order	Nautilida

Genome Traits	Expected	Observed
Haploid size (bp)	817,699,215	949,282,771
Haploid Number	6 (source: ancestor)	26
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 7.7.Q50

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid Number is different from Expected
- . BUSCO single copy value is less than 90% for collapsed
- . Not 90% of assembly in chromosomes for collapsed

Curator notes

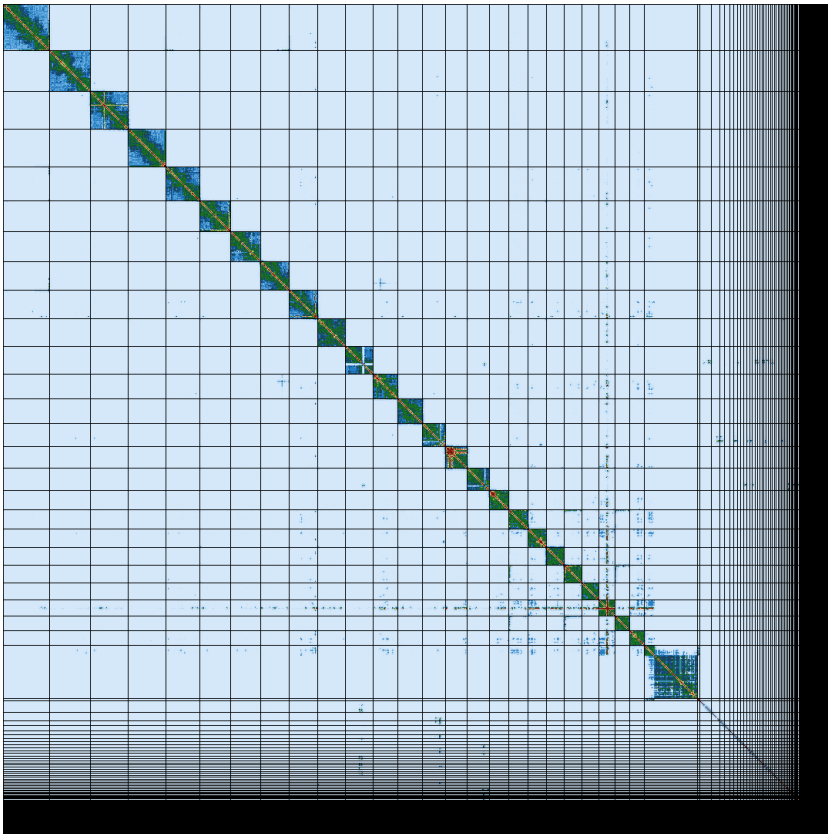
. Interventions/Gb: 41
. Contamination notes: ""
. Other observations: "The assembly of Nautilus macromphalus (xcNauMacr1) is based on 117X ONT data and 201X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 22 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 5.948 Mb (with the largest being 4.224 Mb). Additionally, 2000 regions totaling 75.218 Mb (with the largest being 1.853 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using OATK. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 7 haplotypic regions and 20 contaminant sequences were removed, totaling 3.769 Mb and 3.780 Mb, respectively (with the largest being 2.401 Mb and 1.200 Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	963,268,815	949,282,771
GC %	38.03	38.01
Gaps/Gbp	18.69	38.98
Total gap bp	1,800	6,100
Scaffolds	2,066	246
Scaffold N50	28,443,176	31,537,200
Scaffold L50	13	12
Scaffold L90	51	38
Contigs	2,084	283
Contig N50	20,503,271	20,336,530
Contig L50	16	16
Contig L90	68	67
QV	49.2814	50.3724
Kmer compl.	91.4622	91.3362
BUSCO sing.	81.6%	81.8%
BUSCO dupl.	0.8%	0.6%
BUSCO frag.	12.0%	12.0%
BUSCO miss.	5.7%	5.7%

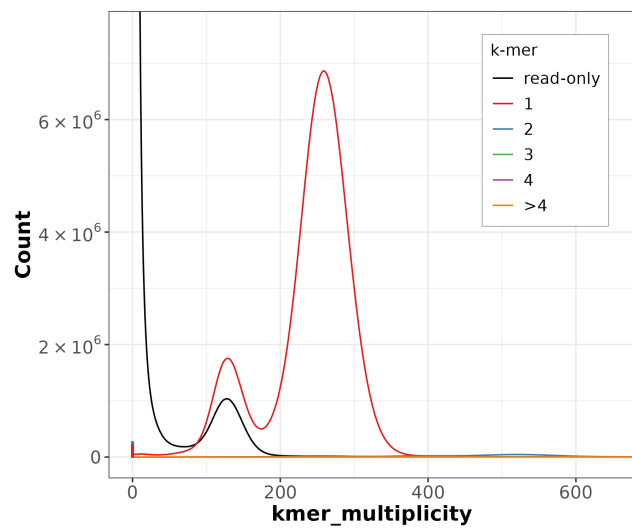
BUSCO: 5.8.2 (euk_genome_met, metaeuk) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

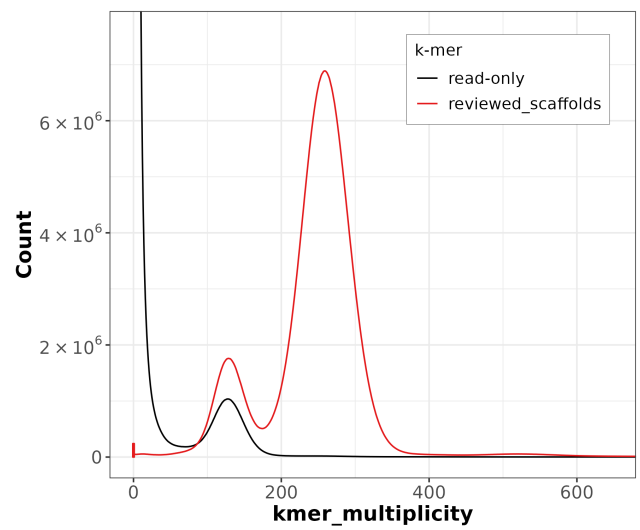


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

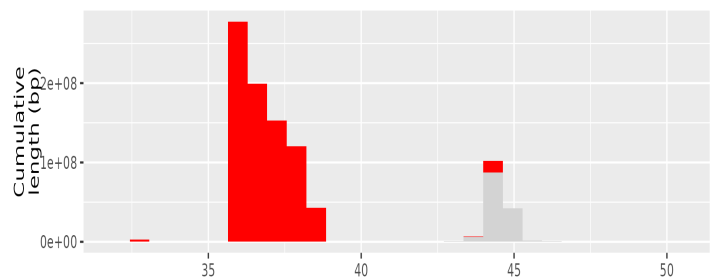


Distribution of k-mer counts per copy numbers found in asm



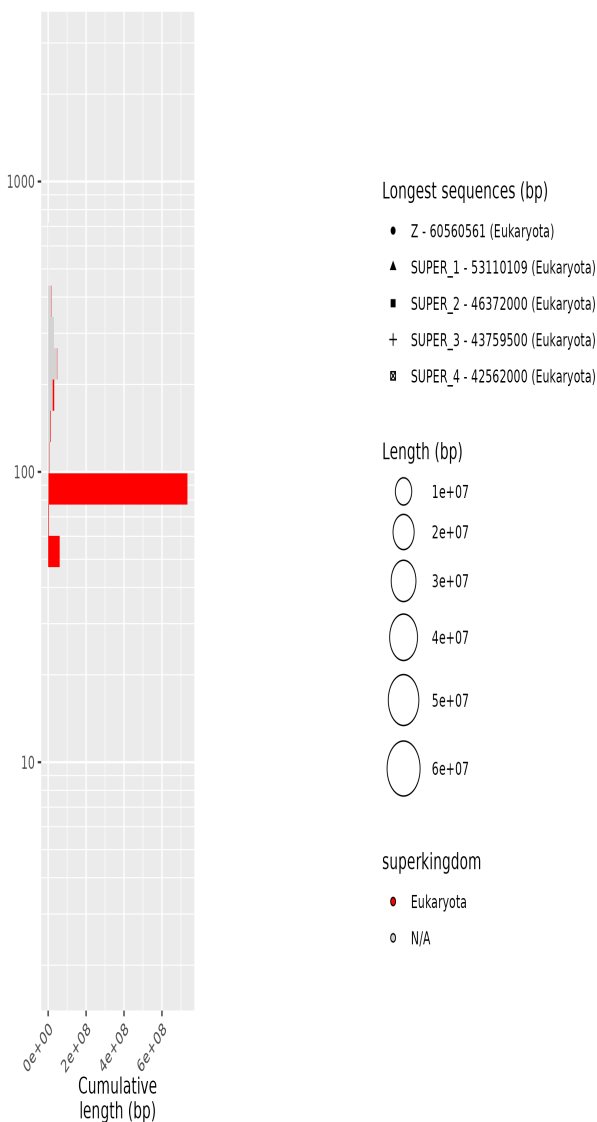
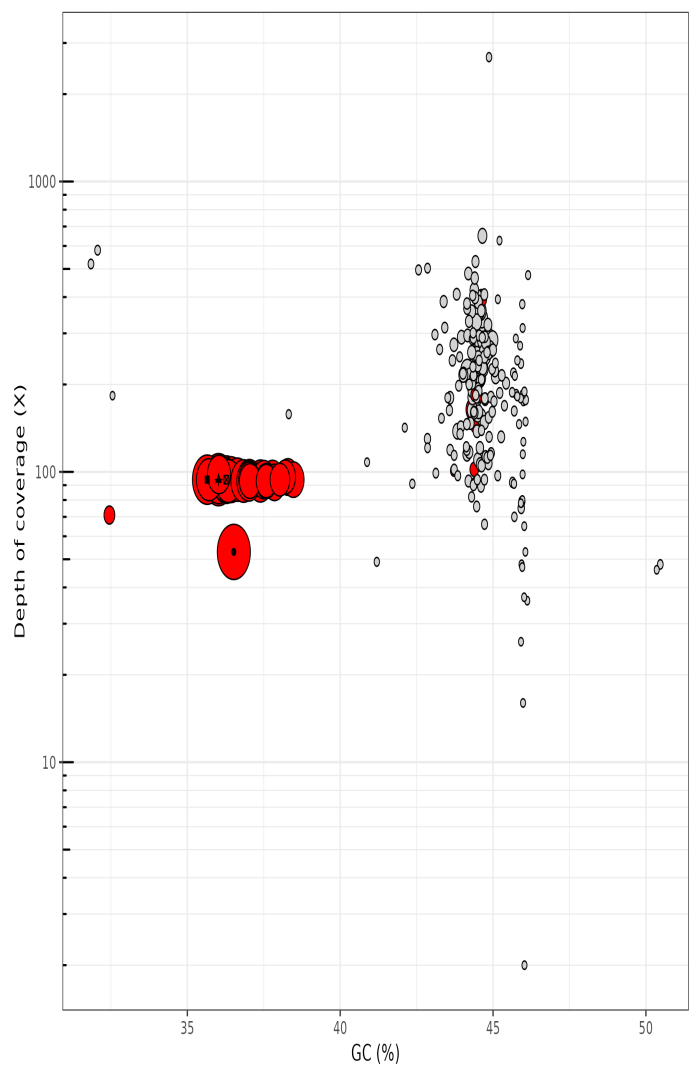
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(9 0X contigs have been hidden)



collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	PACBIO Hifi	Arima
Coverage	117	201

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.19.5-r593
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Curation pipeline

- **PretextMap**
 - |_ *ver*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Benjamin Istace

Affiliation: Genoscope

Date and time: 2025-06-11 17:44:20 CEST